THE EFFECT OF ALTERNATIVE SPLICING ON KEY REGULATORS OF THE INTEGRATED STRESS RESPONSE

Mohammed Alzahrani

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology,
Indiana University

August 2016

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

_____

Ronald C. Wek, Ph.D., Chair

_____

Mark G. Goebl, Ph.D.

_____

Amber L. Mosley, Ph.D.

Mohammed Alzahrani

THE EFFECT OF ALTERNATIVE SPLICING ON KEY REGULATORS OF THE INTEGRATED STRESS RESPONSE

The protein kinase General control non-derepressible-2 (GCN2) is a key regulator of the Integrated stress response that responds to various stress signals, including nutritional deprivation. As a result of high levels of uncharged tRNAs during amino acid depletion, GCN2 phosphorylates serine-51 of the α subunit of eukaryotic initiation factor-2 (eIF2), a translation factor that delivers initiator tRNA to ribosomes. Phosphorylation of eIF2α inhibits general translation, which conserves energy and nutrients and facilitates reprogramming of gene expression for remediation of stress damage. Phosphorylation of eIF2α also directs preferential translation of specific transcription factors, such as *ATF4*. ATF4 reprograms gene expression to alleviate stress damage; however, under chronic stress, ATF4 directs the transcriptional expression of *CHOP*, which can trigger apoptosis. Because multiple stresses can induce eIF2α phosphorylation and translational control in mammals, this pathway is referred to as the Integrated stress response.

GCN2 and *CHOP* are subject to alternative splicing that results in multiple transcripts that differ in the 5'-end of the gene transcripts. However, the effect of the different *GCN2* and *CHOP* isoforms on their function and regulation have not been investigated. Our data suggests that *GCN2* is alternatively spliced into five different transcripts and the beta isoform of *GCN2* is most abundant. Also alternative splicing of *CHOP* creates two *CHOP* transcripts with different 5'-leaders encoding inhibitory upstream open reading frames that are critical for translational control of *CHOP* during stress. This study suggests that alternative splicing can play an integral role in the implementation and regulation of key factors in the Integrated stress response.

Ronald C. Wek, Ph.D., Chair

Table of Contents

List of Tables

List of Figures

List of Abbreviations

αGCN2:	alpha isoform of GCN2
ATF4:	activating transcriptional factor-4
Bcl-2:	B-cell lymphoma-2
βGCN2:	beta isoform of GCN2
BiM:	Bcl-2 interacting mediator of cell death
bp:	base pairs
CDS:	coding sequence
C/EBP:	CCAAT/enhancer-binding protein
CHOP:	CCAAT/enhancer-binding protein-homologous protein
CHOP v1:	CHOP variant-1
CHOP v2:	CHOP variant-2
CHOP v1-luc:	plasmid encoding CHOP v1 uORF upstream luc
CHOP v2-luc:	plasmid encoding CHOP v2 uORF upstream luc
cKD:	catalytic protein kinase domain
Ct:	cycle threshold
CTD:	carboxyl-terminal domain
δGCN2:	delta isoform of GCN2
DR5:	death receptor-5
εGCN2:	epsilon variant of GCN2
eIF2α:	alpha subunit of eukaryotic initiation factor-2
eIF1:	eukaryotic initiation factor-1
eIF2B:	eukaryotic initiation factor-2B
eIF3:	eukaryotic initiation factor-3
ER:	endoplasmic reticulum
EST:	expressed sequence tag
GCN1:	general control non-derepressible-1
GCN2 or EIF2AK4:	general control non-derepressible-2 or eukaryotic initiation factor-2 alpha kinase-4
GCN20:	general control non-derepressible-20

GDP:	guanosine diphosphate
γGCN2:	gamma transcript of GCN2
GTP:	guanosine triphosphate
HisRS:	histidyl-transfer RNA synthetase-like domain
HRI or EIF2AK1:	heme-regulated inhibitor kinase or eukaryotic initiation factor-2 alpha kinase-1
IDT:	Integrated DNA Technologies, Biotechnology company
IP:	intraperitoneal route of injections
ISR:	Integrated stress response
kDa:	kilo Dalton
luc:	firefly luciferase
MEF:	wild type-mouse embryonic fibroblast
Met-tRNA$_i^{Met}$:	initiator methionyl transfer-ribonucleic acid
min:	minutes
NCBI:	National Center of Biotechnology Information
NTD:	amino-terminal domain
OH:	hydroxyl group
PBS:	phosphate-buffered saline
PERK or EIF2AK3:	PKR-like endoplasmic reticulum kinase or eukaryotic initiation factor-2 alpha kinase-3
pKD:	partial protein kinase domain
PKR or EIF2AK2:	protein kinase R or eukaryotic initiation factor-2 alpha kinase-2
pmol:	picomole
PVOD:	pulmonary veno-occlusive disease
5'RACE:	rapid amplification cDNA ends
RCF:	relative centrifugal force
RT:	room temperature
SD:	standard deviation

sec:	seconds
snRNA:	small nuclear RNA
snRNP:	small nuclear ribonucleoprotein
Tg:	an ER stress agent called thapsigargin
Thr:	threonine residue
TK:	thymidine kinase promoter
Tm:	an ER stress agent called tunicamycin
UCSC:	the genome database of University of California, Santa Cruz
U2AF:	U2 auxiliary factor
uORF:	upstream open reading frame
UTR:	untranslated region
UV:	ultraviolet irradiation

Introduction

Regulation of the initiation phase of protein synthesis by eif2α phosphorylation

In response to environmental stresses, cells trigger repression of translation to conserve energy and resources and facilitate reconfiguration of gene expression to better alleviate cell damage. A major regulatory mechanism underlying translational control involves phosphorylation of the serine-51 residue in the alpha subunit of eukaryotic initiation factor-2 (eIF2α). The eIF2 bound to a guanosine triphosphate (GTP) recruits an initiator methionyl transfer-ribonucleic acid (Met-tRNA$_i^{Met}$) to the 40S ribosome. In addition to other eukaryotic initiation factors such as eIF1 and eIF3, this complex constitutes the preinitiation 43S ribosome. The 43S ribosome associates with a cap-binding complex, referred to as eIF4F that is associated with 7-methyl guanosine constituting the 5'- “cap” of mRNAs. The 43S ribosomal complex then scans processively in a 5'- to 3'- direction along the mRNA in search of an appropriate initiation codon. Once the initiation codon, typically an AUG, is recognized via base pairing with the anticodon of the initiator tRNA charged with methionine, the 60S ribosome and the 40S ribosome are joined to form the competent 80S initiation complex and begin the elongation phase of translation.

During the initiation phase of translation, the GTP bound to eIF2α is hydrolyzed to guanosine diphosphate (GDP), and the association of the small and large ribosomal subunits is preceded by with the release of eIF2-GDP. The eIF2-GDP must be recycled to the active GTP bound form to begin subsequent rounds of translation initiation. This eIF2-GDP to eIF2-GTP exchange is facilitated by the eukaryotic initiation factor-2B (eIF2B). However, the eIF2B-mediated recycling process is competitively inhibited by eIF2α phosphorylation, which occurs during various cellular stresses (Baird & Wek, 2012; Hinnebusch, 2005; Jackson, Hellen, & Pestova, 2010). The consequence of lowered eIF2-GTP is reduced delivery of the initiator tRNA to the 40S ribosome and sharply lowered global translation.

Family of eIF2α kinases function in the Integrated stress response

In mammalian cells, four different eIF2α kinases sense cellular stresses and trigger global repression of translation initiation (Figure1). Family members include heme-regulated inhibitor kinase (HRI or EIF2AK1) that is induced by heme-deprivation, heat shock, or oxidative damage (Shao et al., 2001); Protein kinase R (PKR or EIF2AK2) which functions in an anti-viral induced by interferons; PKR-like endoplasmic reticulum kinase (PERK or EIF2AK3) that is stimulated by accumulation of misfolded proteins residing in the endoplasmic reticulum (ER); General control non-derepressible-2 (GCN2 or EIF2AK4) that is upregulated in nutrient-deprived cells (Figure 1) (Baird & Wek, 2012). While activation of each eIF2α kinase leads to repression of global translation initiation, eIF2α phosphorylation also facilitates preferential translation of key stress-related mRNAs. For example, translation of the *Activating transcriptional factor-4* (*ATF4*) is induced by phosphorylation of eIF2α during cellular stress. Selective translation of *ATF4* has been proposed to occur by a mechanism involving two upstream open reading frames (uORF) situated in the 5'-end of *ATF4* mRNAs (Harding et al., 2000; Vattem & Wek, 2004). In this model of delayed translation reinitiation, following translation of the 5'-proximal uORF1 there is retention of the 40S ribosome on the mRNA, which continues scanning 5' to 3' in search of the next open reading frame. In non-stressed conditions and low levels of eIF2α phosphorylation, ribosomes are proposed to rapidly reacquire eIF2-GTP-Met-$tRNA_i^{Met}$ and translate the next open reading frame, uORF2. The uORF2 overlaps out-of-frame with the *ATF4* coding sequence (CDS) and consequently there is minimal expression of ATF4 protein. During stressed conditions and elevated levels of eIF2α phosphorylation that reduce eIF2-GTP levels, the reinitiating ribosomes require an extended time to reacquire the limited eIF2-GTP-Met-$tRNA_i^{Met}$. As a consequence of this delay, the 40S ribosome scans past the uORF2 initiation codon before acquisition of eIF2-GTP-Met-$tRNA_i^{Met}$ that allows for reinitiation of translation at the *ATF4* CDS (Vattem & Wek, 2004). The resulting increase in levels of ATF4 protein leads to increased transcription of target genes involved in amino acid metabolism, nutrient uptake, and

suppression of oxidation. Because multiple different stresses induce eIF2α phosphorylation and preferential translation of *ATF4*, this pathway has been referred to as the Integrated stress response (ISR) (Harding et al., 2003).

Dysfunction in the ISR leads to disease

Genetic changes in the eIF2α kinases have been linked to pathologies. For instance, *HRI* knockouts in mice are associated with hematological disorders such as iron deficiency anemia, erythropoietic protoporphyria and β- thalassemia (Chen, 2007). Loss of the antiviral activity of *PKR* can be demonstrated in rift valley fever patients whose clinical signs ranging from an elevated body temperature to retinitis, hepatic inflammation, kidney dysfunction, encephalitis, lethal hemorrhagic and death (Habjan et al., 2009). Another *PKR*-associated disease is Fanconi anemia complementation group C that is characterized by bone marrow failure, congenital defects, and tumor susceptibility (Pang, Christianson, Keeble, Koretsky, & Bagby, 2002). In humans, loss of *PERK* leads to Wolcott-Rallison syndrome, which features insulin-dependent diabetes, osteoporosis, and growth retardation (Delepine et al., 2000). A recent study has reported that *GCN2* mutations are the underlying cause of the pulmonary veno-occlusive disease (PVOD) (Eyries et al., 2014). PVOD is a rare form of pulmonary hypertension and characterized by narrow pulmonary veins and venules because of increased cell proliferation. In addition, *GCN2* is involved in cell lipid metabolism (Guo & Cavener, 2007), learning and memory, the survival pathway of tumor cells (Costa-Mattioli et al., 2005), immune response and DNA repair under the influence of ultraviolet irradiation (UV) (Jiang & Wek, 2005). Finally, knocking out *ATF4* in mice, the common target of these eIF2α kinases, results in abnormal lens morphogenesis, anemia, delayed growth and bone development (Masuoka & Townes, 2002; Tanaka et al., 1998).

GCN2 is a multidomain protein kinase activated by diverse stresses

The protein kinase GCN2 is composed of multiple domains that are shared among diverse species, including human, mice, and *Drosophila*

(Berlanga, Santoyo, & De Haro, 1999; Sood, Porter, Olsen, Cavener, & Wek, 2000). GCN2 domains include the amino-terminal domain (NTD), the partial protein kinase domain (pKD), the catalytic protein kinase domain (cKD), the histidyl-transfer RNA synthetase (HisRS)-like domain, and the carboxyl-terminal domain (CTD) (Figure 2) (Hinnebusch, 2005; Qiu, Dong, Hu, Francklyn, & Hinnebusch, 2001; Wek, Jackson, & Hinnebusch, 1989). Each of these domains is critical for activation of GCN2 during nutrient deprivation. As will be expanded upon below, *GCN2* has a number of exons and its alternative splicing can affect the expression of GCN2 protein with different domain arrangements.

The activity of GCN2 is stimulated by mechanisms contributing to the accumulation of non-aminoacylated tRNAs. Stressed conditions activating GCN2 include reduced amino acid availability and high salinity (Zaborske et al., 2009), glucose starvation (Yang, Wek, & Wek, 2000), and UV irradiation that have been suggested to increase the level of deacylated tRNAs (Kwon et al., 2011). Activation of GCN2 by uncharged tRNAs is driven by the HisRS-like domain; binding of uncharged tRNAs to HisRS domains overcomes the inhibitory effect of CTD and allows GCN2 activation and dimerization (Lageix, Zhang, Rothenburg, & Hinnebusch, 2015). To reach the maximum activity of GCN2, GCN2 undergoes auto-phosphorylation of the two conserved threonine residues, Thr-882 and -886 in budding yeast and Thr-898 and -903 in mice, located within the activation loop of cKD (Castilho et al., 2014; Romano et al., 1998). The N-terminal portion of GCN2 has been proposed to be the docking site of the GCN1-GCN20 complex, and deletion of this region in GCN2 thwarts activation of GCN2 by nutrient stresses. It is suggested that the GCN1-GCN20 complex is associated with ribosomes and may contribute to GCN2 recognition and binding to uncharged tRNAs. The pKD lacks critical residues for the kinase activity, but allosterically enhances the activation of the neighboring cKD (Garcia-Barrio, Dong, Ufano, & Hinnebusch, 2000; Lageix, Rothenburg, Dever, & Hinnebusch, 2014; Sood et al., 2000).

Multiple ISR genes are subject to preferential translation and control cell fate during stress

As a consequence of diverse stress signals, there is induction of eIF2α phosphorylation that serves to inhibit overall protein production coincident with preferential translation of *ATF4* (Baird & Wek, 2012). Induced ATF4 then modulates the transcription of a collection of ISR genes involved in restoring cellular homeostasis or alternatively, reducing cellular viability. For example, CCAAT/enhancer-binding protein (C/EBP)-homologous protein (*CHOP*) encodes the carboxyl-terminal bZIP domain and drives an apoptotic pathway during chronic periods of stress (Palam, Baird, & Wek, 2011). *CHOP* was originally discovered as a member of the C/EBP family that inhibits other C/EBP-mediated gene transcription by dimerization (Ron & Habener, 1992). Thus, elevated levels of CHOP protein may abolish the effect of anti-apoptotic proteins such as B-cell lymphoma-2 (Bcl-2) favoring cell cycle arrest and apoptosis (McCullough, Martindale, Klotz, Aw, & Holbrook, 2001).

CHOP also upregulates apoptotic genes such as Bcl-2 interacting mediator of cell death (*BiM*) and death receptor-5 (*DR5*) (Puthalakath et al., 2007; Yamaguchi & Wang, 2004). The proapoptotic BiM protein belongs to the BCL-2 family and is required for mitochondrial permeabilization and subsequently apoptosis. Loss of *BiM* in mice leads to resistance to ER stress-induced apoptosis of renal epithelium (Tabas & Ron, 2011). Another CHOP-targeted gene is *DR5* that triggers the extrinsic pathway of apoptosis when DR5 is bound to a ligand. Ligand binding to DR5 activates caspase-8, which amplifies apoptotic signaling by cleaving and activating other procaspases, such as caspases-3, -6, and -7, leading to cell death (McIlwain, Berger, & Mak, 2013). DR5 significantly mediates apoptotic induction of human tumor cells treated with ER stress agents such as thapsigargin (Tg) (Xu, Su, & Liu, 2012; Yamaguchi & Wang, 2004). In addition to induced transcriptional expression, *CHOP* mRNA is preferentially translated by eIF2α phosphorylation. The mechanism of induced *CHOP* translation involves a single uORF that serves as a potent inhibitor of *CHOP* translation during non-stressed conditions; however, upon stress and eIF2α

phosphorylation, scanning ribosomes can bypass the *CHOP* uORF and instead initiate translation at the downstream coding region of the *CHOP* mRNA (Palam et al., 2011).

Alternative splicing and gene expression

Alternative splicing is a post-transcriptional mechanism responsible for mRNA diversity through various splicing patterns, e.g. exon skipping, alternative 3'- and 5'-splice sites, mutually exclusive exons, and intron retention. Alternative splicing dictates which portions of the genetic sequences are included in and/or excluded from encoded mRNAs. These splicing changes can alter the coding sequence and thus modify the domain arrangement of the synthesized protein. Alternative splicing can also change the 5'- or 3'-untranslated regions (UTR) of the transcript that can influence mRNA stability or levels of translation (Kalsotra & Cooper, 2011). Differences in the 5'-UTRs of the mRNAs can also result from differences in transcriptional start sites, and selection of polyadenylation recognition sequences that define the site for pre-mRNA cleavage and polyadenylation. These differences in pre-mRNA processing can alter the presence of uORFs in the 5'-end of the mRNAs or alternatively modify protein or miRNA binding sequences in the 3'-portion of the transcript that direct mRNA turnover (Stamm et al., 2005).

Splicing of the pre-mRNA is mediated by the spliceosome. Spliceosomes are central for identifying conserved sequences located at the exon and intron junctions of introns, and the spliceosome in conjunction with associated factors can recognize splicing enhancers and silencers that determine which introns are removed or retained. The spliceosome is a protein complex encompassing small nuclear RNAs (snRNA) and snRNA-associated proteins to form small nuclear ribonucleoproteins (snRNP) (Goren et al., 2006). The protein complex is assembled by the binding of U1 snRNP to the conserved AG dinucleotide at the 5'-splice site of the nascent transcript via base pairing. Splicing factor U1 and U2 auxiliary factor (U2AF) are attached to consensus sequences constituting the 3'-splice site portion; these sequences represent the branch point, polypyrmidine

tract, and terminal AG dinucleotide. Together, U1 and U2AF comprise the early or commitment complex that is joined by U2 snRNP at the branch point forming the A complex. The A complex recruits a trimeric complex containing U4, U5, and U6 snRNPs to create the B complex. This is followed by the formation of the C complex due to conformational changes in which U1 snRNP is replaced with U6 snRNP at the 5'-splice site and U1 and U4 snRNPs are released from the complex (Black, 2003).

Chemically, the spliceosome complex performs two transesterification reactions; the first reaction commences by a cleavage between the 5’-exon junction and intron caused by the nucleophilic attack of the 2'-hydroxyl group of a conserved adenosine at the branch point on the phosphate at the 5'-splice site. This reaction leads to formation of a phosphodiester linkage between the 5'-end of the intron and 2’-OH of the branch point in a lariat configuration. The second transesterification reaction occurs when the free reactive 3'-OH at the 3'-end of the released exon attacks the phosphodiester bond connecting the 3'-end of the intron to the 5'-end of the downstream exon. The two exons are ligated and an lariat intron is excised (Staley & Guthrie, 1998).

Alternative splicing of GCN2 mRNA

GCN2 is alternatively spliced into at least three different transcripts, designated alpha (αGCN2), beta (β*GCN2*), and gamma (γ*GCN2*) that effect the coding sequence and expressed N-terminal portion of GCN2. These *GCN2* splicing variants were first identified by using rapid amplification cDNA ends (5’-RACE) and sequencing methods (Sood et al., 2000). The *GCN2* mRNA is ubiquitously expressed and abundant in brain and liver tissues (Berlanga et al., 1999); however, alpha and gamma isoforms of *GCN2* were restricted to brain and testis (Sood et al., 2000). Therefore, there can be variations in the NTD and pKD regions of GCN2 that are generated by these splicing events.

Analysis of *CHOP* expressed sequence tags (EST) suggests that *CHOP* is also subject to alternative splicing, with two isoforms based on the inclusion or exclusion of the second exon. Although *CHOP* expression is widely distributed

among tissues, the levels of CHOP protein are low during non-stressed conditions. By contrast, synthesis of CHOP protein is induced during stress and this transcriptional factor is then translocated to the nucleus to mediate transcriptional regulation (Tabas & Ron, 2011).

Questions addressed in this thesis research

In this study, we investigated the effect of alternative splicing on the key ISR regulators, *GCN2* and *CHOP*. Although *GCN2* and *CHOP* isoforms have been suggested, their differential expression among tissues and functional roles are still obscure. We addressed several questions. First, what are the *GCN2* and *CHOP* splicing variants? Second, are there differences in the regulation of CHOP translation by the alternative splicing in the 5'-portion of the *CHOP* mRNA during non-stressed and stressed conditions? Finally, are there differences in the expression of *GCN2* and *CHOP* transcripts among different mouse tissues? Addressing these fundamental questions will provide new insight into the roles of alternative splicing in the regulation and functions of ISR.

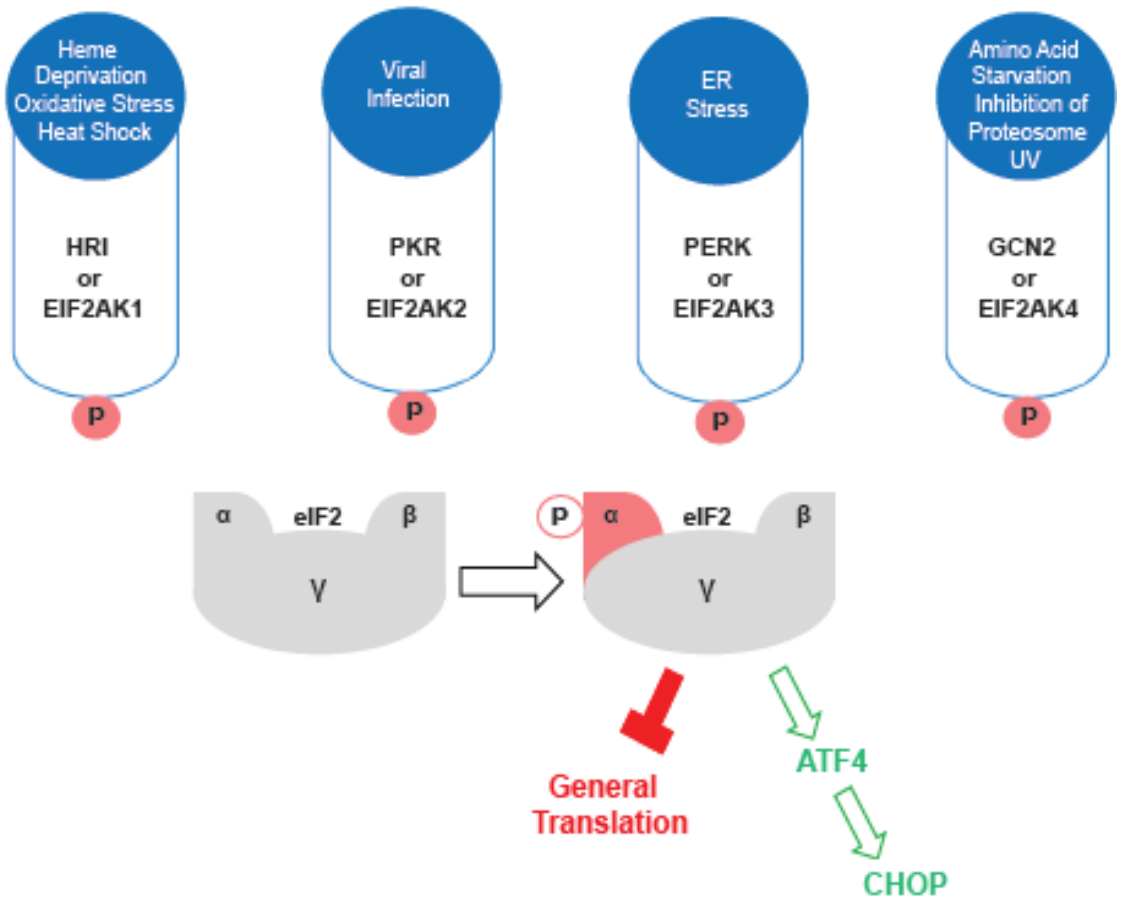

Figure 1: The Integrated stress response results in inhibition of overall protein alongside of preferential translation of some stress-related gene transcripts. The eukaryotic initiation factor-2 alpha (eIF2α) kinases are the Heme-regulated inhibitor (HRI), Protein kinase R (PKR), PKR-like endoplasmic reticulum kinase (PERK), and General control non-derepressible-2 (GCN2). Each of these protein kinases is activated by different stress signals. Phosphorylation of eIF2α inhibits general protein production, but selectively induces the translation of specific mRNAs, such as the *activating transcriptional factor-4* (*ATF4*). ATF4 activates the transcription of other stress-related genes to ameliorate the stress response or induce apoptosis by C/EBP homologous protein (CHOP). ER, endoplasmic reticulum; UV, ultraviolet irradiation; and P, phosphorylation.

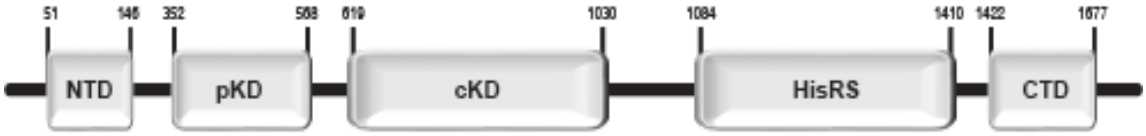

Figure 2: Protein domains of the protein kinase GCN2. GCN2 consists of multiple domains that interact with each other and other proteins for regulation of GCN2 activity. GCN2 domains include the N-terminal domain (NTD), the partial kinase domain (pKD), the catalytic kinase domain (cKD), the histidyl-transfer RNA synthetase-like domain (HisRS), and the C-terminal domain (CTD). The length of each domain is represented in amino acid residues, and domains were demarcated by using the software program Pfam.

Methods and Materials

Identification of GCN2 variants in the genome

The sequence of *GCN2* isoforms were derived from the genome database created by University of California, Santa Cruz (UCSC). We used a sequence alignment program called CLUSTALW to identify the unique sequence of each splicing isoform based on the sequence differences compared to the sequence of β*GCN2*. The unique sequence of each isoform was used as a query in the BLAST program, provided by National Center of Biotechnology Information (NCBI), to search the mouse database for ESTs. Thus, we addressed whether the *GCN2* splicing variant is likely to be expressed.

Tissue homogenization and RNA isolation

The protocols of animal study were approved by the Institutional Animal Care and Use Committees at Rutgers University and the Indiana University School of Medicine. Different mouse organs including whole brain, liver, and testis were extracted from Cre-mediated red fluorescent reporter transgenic mice with C57BL/6 background (Banks et al., 2011). The other liver tissues, used for CHOP PCR, were also derived from Cre-negative mice born after breeding between C57BL/6 mice carrying floxed alleles of *ATF4* and C57BL/6 mice carrying one allele of Cre recombinase driven by an albumin promoter. The Cre-negative mice were intraperitoneally (IP) injected with 1 mg of tunicamycin (Tm)/ kg body weight or 0.3% DMSO in phosphate-buffered saline (PBS) as a control, i.e. non-treated mice. These mice were sacrificed by decapitation after 6-hour-treatment with Tm (Fusakio et al., 2016). The tissues were rinsed with cold PBS and immediately frozen in liquid nitrogen after being covered with aluminum foil. By using clear mortar and pestle placed on dried ice, each organ was homogenized in the presence of liquid nitrogen and the ground tissues were frozen at - 80 °C. 2-30 mg of ground tissues were used for RNA purification. Addition of 1 ml of TRIzol reagent (Ambion, Life Technologies) was used to extract RNA after being released from disrupted cells by a mechanical force.

After 5 minutes (min) of incubation at room temperature (RT), 200 µl of chloroform was added and solution was incubated for 3 min at RT. The solution was clarified by centrifugation at 12,000 relative centrifugal force (RCF) for 15 min at 4 °C, and supernatant was collected. To precipitate RNA, 500 µl of isopropyl alcohol was added to supernatant and incubated for 10 min at RT. Following centrifugation at 12,000 RCF for 10 min at 4°C, pellets were collected and washed with 1 ml of 75 % ethanol. Ethanol was removed following an additional centrifugation at 7,500 RCF for 5 min at 4 °C, and pellets were placed on ice.

RNA samples were further purified by using the RNeasy® Plus Mini Kit (50) (QIAGEN); pellets were dissolved in 100 µl of RNase free water. 350 µl of freshly made solution of RLT buffer and β-mercaptoethanol was added to the dissolved pellets. After mixing the solution by a vortex, 250 µl of absolute ethanol was added creating small precipitated particles. The total volume, 700 µl, was transferred to an RNase easy column that was subjected to centrifugation at 8,000 RCF for 15 seconds (sec). The columns were washed with 500 µl of an ethanol-added RPE buffer after being centrifugation at 8000 RCF for 15 sec. A final centrifugation was carried out for the columns at 12,000 RCF for 1 min. The final step in the purification involved addition of 35 µl RNase free water to columns, which were incubated for 5 min at RT before the RNA was eluted at 14,000 RCF for 1 min. The RNA concentrations were measured by using a Nano-Drop spectrophotometer.

Reverse transcription and PCR

Purified total RNA was reverse transcribed to produce cDNAs using the Taqman regent and Multiscribe Reverse Transcriptase (Applied Biosystems). In brief, 2 µg of purified RNA was diluted in 19 µl of RNase free water was added to 30 µl of the Taqman reagent mix containing 10X Taqman reagent, 25 mM of magnesium chloride, 2.5 mM deoxynucleotide mix, 1.25 mM random hexamers, and 20 Units RNase inhibitor. The addition of 1 µl of Multiscribe Reverse Transcriptase brought the volume to 50 µl. The reverse transcription was

performed for 48 minutes at 48°C after activating the enzyme at 25 °C for 10 min. To inactivate the enzyme and the reverse transcription, the reaction was exposed to high temperature, 95 °C for 5 min, and the concentration of cDNA was measured by the Nano-Drop spectrophotometer.

To detect *GCN2* transcripts, we designed multiple sets of primers that were specific to each isoform of *GCN2* (Table 1). PCR reactions contained 12.5 µl of R-Taq Master Mix (Bullseye), 1 µl of the specific forward and reverse primers corresponding to each isoform, 1 µl of a template DNA (0.1-1 µg), and 9.5 µl of distilled water. The PCR reaction was programmed for 35 cycles in which each cycle featured three steps; the denaturation step was set at 94 °C for 30 sec; the annealing step was set at 60 °C for 30 sec; the elongation step was set at 72 °C for 15 sec. An additional 10 min-elongation step was assigned before inactivating the polymerase at 4 °C. The amplicons were separated by electrophoresis using an electrical field of 104-116 Volts applied on 1.5 % of an agarose gel. Adding ethidium bromide to the gel helped visualize the DNA bands separated by gel electrophoresis and UV light.

Similarly, we relied on PCR to detect *CHOP* variants using specific primers designed upon the alternative exon of *CHOP* (Table 2). The PCR reaction was set up as described earlier with similar programming steps, except for the number of PCR cycles and the annealing temperature; PCR cycles were reduced to 20 cycles, and the annealing step was set at 72 °C for 30 sec.

Quantitative PCR analysis

To estimate the abundance of the β*GCN2* within different mouse tissues, we used the real time-polymerase chain reaction (RT-PCR). First, we amplified a PCR product containing only β*GCN2*. Then, we made serial dilutions of the concentrated β*GCN2* ranging from 1 picomole (pmol) to $1x10^{-08}$ pmol. To build a stable and reproducible standard curve for β*GCN2*, 5 µl of each dilution of β*GCN2* was plated in triplicates and mixed with 15 µl of the SYBR mix including 1 µl of 5 µM forward and reverse β*GCN2* primers, 3 µl of distilled water, and 10 µl of SensiMix™ SYBR® No-ROX Kit, BIOLINE. Next, we performed RT-PCR on

different mouse tissues using the specific β*GCN2* primers. This RT-PCR generated double stranded-DNAs of β*GCN2* that were detected by the fluorescent (SYBR Green) dye. Results of RT-PCR were represented as the mean of cycle threshold (Ct) values of the three independent samples, i.e. triplicates, plus or minus the standard deviation (± SD). The abundance of β*GCN2* in different samples of mouse tissues was deduced by comparing the Ct values of β*GCN2* in each sample to that of the standard curve.

Plasmids and dual luciferase reporter assays

Four different plasmids were utilized in this study; two pcDNA3.1-derived plasmids expressing the full sequence of either *CHOP* variant-1 (CHOP v1) or *CHOP* variant-2 (CHOP v2) were used to check the specificity of *CHOP* primers. As a PCR template for *CHOP* primers, we first used the full *CHOP* v1 cDNA plasmid (*CHOP* v1 plasmid) encoding 1 kilobase of *CHOP* promoter upstream of the 5'-leader and CDS of *CHOP* v1 (Young, Palam, Wu, Sachs, & Wek, 2016). To construct the full version of *CHOP* v2 cDNA (*CHOP* v2 plasmid), the *CHOP* v1 plasmid was digested with HindIII and BsmBI restriction enzymes (New England BioLabs) to remove the 5'-leader of *CHOP* v1 and this portion was replaced with the analogous DNA portion encoding the 5'-leader of *CHOP* v2 which was synthesized by Integrated DNA Technologies (IDT). The resulting *CHOP* v2 plasmid was then utilized as a PCR template to confirm that the *CHOP* primers in Table 2 are specific to their corresponding transcripts.

The other two pGL3-derived plasmids encoding either *CHOP* v1 uORF (CHOP v1-luc) or *CHOP* v2 uORF (CHOP v2-luc), each with the start codon of *CHOP* CDS fused in-frame to the coding sequence of firefly luciferase. Transcription of the CHOP-Luc fusion genes were expressed constitutively from a *thymidine kinase* promoter (TK) as previously described (Palam et al., 2011; Young et al., 2016). To transfect MEF cells with CHOP v1-luc or CHOP v2-luc constructs, we first cultured MEF cells for 24 hours and then split cells into two six-well plates designated as CHOP v1 plate and CHOP v2 plate. MEF cells in each plate were then transiently co-transfected with either CHOP v1-luc or

CHOP v2-luc plasmids and a *Renilla* luciferase construct based on the protocol of FuGENE 6 transfection reagent (Roche Applied Science). After incubating the cells for 24 hours, triplicate wells within each plate were treated with 0.1 μM of Tg or no ER stress agent and followed by an additional 12 hour incubation (Palam et al., 2011). MEF cells were lysed by adding 500 μl of 1X passive lysis solution (Promega) to each well followed by shaking for 15 min. Following centrifugation at 13,000 RCF for 2 min at 4°C, 20 μl of the supernatant was used to conduct the dual-luciferase assay. The luciferase activity was represented as relative light units of the ratio of firefly to *Renilla* luciferase units (Palam et al., 2011). Relative values from the triplicate wells treated with thapsigargin for 12 hours were averaged and normalized to the mean value of the non-treated triplicate wells. Relative values are shown as CHOP v1 and CHOP v2 histograms with the SD indicated.

Statistical analysis

Relative values of stressed and non-stressed CHOP v1 and CHOP v2 with ± SD are derived from biological triplicates. By using the two-tailed student's t-test, P values were calculated to indicate the significance of the test. If P values were 5% or less, the test was considered statistically significant and indicated by "*".

Table 1: Sequences of primer sets used for the identification of *GCN2* isoforms.

Sets	Forward primer sequence	Reverse primer sequence	Amplicons
Pair 1	5'-CATCCGCTTGTATGAG CAGCTA-3'	5'-CTGACGTTCTCGCCT GGA-3'	α*GCN2*
Pair 2	5'-CCTACATACCCAGATG TAGTTCCCGAA-3'	5'-CCAGCATTTCTTCATG GAAAGACTTTGGTG-3'	β*GCN2*
Pair 3	5'-GCGGGCTCTTCTAGT TCCCGAAAT-3'	5'-GCCTTTCCAGCATTTC TTCATGGAAAGAC-3'	γ*GCN2*
Pair 4	5'-CCAAGACTTCCTGAA GAATCCTAGGAGAC-3'	5'-TGCACTTGTGATTCAC AAGTGAAAAATCCTC-3'	δ*GCN2*
Pair 5	5'-CCCACCTACATACCC AGATGTGTGATG-3'	5'-CCTCTTTCCTTCTCTG AATCTCGTGGAG-3'	ε*GCN2*

Table 2: Sequences of primer sets used for the identification of *CHOP* isoforms.

Sets	Forward primer sequence	Reverse primer sequence	Amplicons
Pair 1	5'-CCTGAGGAGAGAGTG TTCCAGAAGG-3'	5'-CGTTTCCTGGGGATG AGATATAGGTGC-3'	*CHOP* v1
Pair 2	5'-CTGAGGAGAGAGAAC CTGGTCCA-3'	5'-ACGCAGGGTCAAGAG TAGTGAAG-3'	*CHOP* v2

Table 3: Sequences of primers used for the identification of firefly luciferase and *β-actin*.

Amplicons	Forward primer sequence	Reverse primer sequence
Firefly luc	5'-CTCACTGAGACTACATCAGC-3'	5'-TCCAGATCCACAACCTTCGC-3'
β-actin	5'-TGTTACCAACTGGGACGACA-3'	5'-GGGGTGTTGAAGGTCTCAA-3'

Results

Characterization of mouse *GCN2* in the genome

There are five putative *GCN2* splicing variants reported in the UC Santa Cruz genome database for mice (GRCh38/hg38). To further address the existence of *GCN2* isoforms in the genome, we searched the EST database of NCBI for sequences specific to each isoform of *GCN2*. In addition to the previous identified splicing variants (Berlanga et al., 1999; Sood et al., 2000), we found an EST for a new isoform, ε*GCN2*, with an accession number of CD355187, which belongs to *Mus musculus* cDNA extracted from whole brain. Together, these ESTs support the idea that there are several *GCN2* isoforms expressed in mice. However, there is no EST specific for the delta isoform of GCN2 (δ*GCN2)*; therefore, we used the δ*GCN2* primers to amplify and detect the δ*GCN2* mRNA in non-treated MEF cells as a proof of δGCN2 existence in the genome (Table 1). After 35 cycles of PCR, we detected low expression of δ*GCN2* in the MEF cells. Next, we cut the δ*GCN2* band at the size of 170 base pairs (bp) in the agarose gel (Figure 3) and extracted δ*GCN2* DNA according to the manufacture's protocol of QIAquick® Gel Extraction Kit (QIAGEN). The extracted δ*GCN2* DNA was sequenced by GENEWIZ and the sequence of the extracted DNA was similar to δ*GCN2* sequence derived from UCSC, suggesting the existence of δ*GCN2* in the genome.

Mouse *GCN2* has five isoforms with different N-termini

There are three reported variants of *GCN2* mRNA, namely alpha, beta and gamma (Sood et al., 2000). In addition to alternative splicing of *GCN2*, the three *GCN2* isoforms are also suggested to arise from alternative transcriptional sites (Sood et al., 2000; Zhang et al., 2002). β*GCN2* represents the full-length version of *GCN2* encoding a large protein of 1648 residues with a molecular weight of 186 kilo Dalton (kDa). The number of exons and sequence of these *GCN2* isoforms were reported (Sood et al., 2000). The five *GCN2* isoforms were identified in non-treated MEF cells using the specific primers for each isoform

(Table 1); first, α*GCN2* was amplified at 214 bp in the agarose gel; secondly, β*GCN2* was detected at 211 bp in the gel; third, γ*GCN2* was observed at 212 bp in the gel; lastly, the new delta and epsilon variants of *GCN2*, δ*GCN2* and ε*GCN2*, were detected with the size of 172 and 214 bp in the agarose gel, respectively. Detection of *GCN2* splicing variants was confirmed by sequencing the *GCN2* bands. In addition, the absence of other bands with larger sizes emphasizes the specificity of *GCN2* primers (Figure 3A).

The sequence differences between *GCN2* isoforms take place within the encoded amino terminal domain (Figure 3). β*GCN2* represents the full-length version *GCN2* encoded in 39 exons, and α*GCN2* lacks the first five exons of β*GCN2* representing the sequence of NTD; the first five exons of α*GCN2* are replaced with an elongated exon as the first and alternative exon of α*GCN2* (Zhang et al., 2002). The gamma transcript contains an alternative exon between the second and third exons of the beta transcript. δ*GCN2* lacks the N-terminal domain and parts of the pKD sequence, but includes an extra exon situated between the ninth and tenth exons of β*GCN2*. In addition, δ*GCN2* transcripts encode the shortest protein of GCN2 because the translation initiation codon is located within the cKD of GCN2. ε*GCN2*, on the other hand, has an additional exon between the first and second exons of β*GCN2*, but ε*GCN2* is missing the third exon of β*GCN2* (Figure 3B).

Expression of β*GCN2* transcripts is higher in mouse testis than brain and liver

β*GCN2* transcripts are expressed in different mouse tissues including whole mouse brain, liver and testis as estimated by RT-PCR. We performed PCR using 0.1 µg of mouse brain, liver, or testis cDNAs as templates to amplify the β*GCN2* DNA at the size of 211 bp. As seen in the PCR result (Figure 4A), the intensity of the β*GCN2* bands derived from testis, reflects how much β*GCN2* RNA is expressed in this tissue. We suggest that β*GCN2* expression is higher in testis than the other tissues. To further support this claim, we developed a standard curve for β*GCN2* to be used as a control for β*GCN2* measurement in

the different mouse tissues. PCR of different cDNA concentrations, 0.2 µg, 0.4 µg, 0.6 µg, 0.8 µg, 1 µg, and 1.2µg, from the whole mouse brain, liver or testis was used to estimate the Ct values for each sample. The Ct values of these samples were then compared to that of the β*GCN2* standard curve to measure the amount of β*GCN2* cDNA prepared from each sample. The β*GCN2* measurement was calculated by plotting the Ct values of each sample in the Y variable in the slope equation of β*GCN2* standard curve (Y = - 3.7254X - 4.1254) (Table 4). Ct values are inversely proportional to the number of β*GCN2* transcripts. Based on RT-PCR analysis, the Ct values of liver samples were lower than of brain samples, indicating a higher amount of β*GCN2* mRNAs in mouse liver tissues. The testis samples showed the lowest Ct values, revealing the highest levels of β*GCN2* transcripts (Figure 4B). Therefore, β*GCN2* mRNAs are abundant in testis, moderate in liver, and lowest in brain tissues.

The two isoforms of mouse *CHOP* are determined by the inclusion or exclusion of the second exon

CHOP encodes a 29-kDa protein composed of 168 amino acids (Ron & Habener, 1992). *CHOP* mRNAs are derived from 4 exons; the first two exons represent the 5'-leader of *CHOP* and the remaining 3 and 4 exons encode the bZIP domain. As a consequence of the alternative splicing at the second exon, the sequence of the 5'-leader of *CHOP* is varied, creating two isoforms-called *CHOP* v1 and *CHOP* v2. *CHOP* v1 mRNA contains 4 exons in which the second exon is included, whereas *CHOP* v2 mRNA lacks the sequence of the alternative exon and thus includes 3 exons. *CHOP* v1 mRNA contains an uORF encoding a 34-residue polypeptide that proceeds the *CHOP* CDS. Due to the removal of the second exon containing a stop codon, the *CHOP* v2 uORF overlaps with the *CHOP* CDS in a different reading frame, encoding a larger polypeptide of 87 residues (Figure 5A) (Palam et al., 2011).

To distinguish between the two *CHOP* isoforms, we designed specific *CHOP* primers based on the presence or absence of the second exon (Table 2). Using the *CHOP* primers with 0.01 µg of either *CHOP* v1 plasmid or *CHOP* v2

plasmid DNA was sufficient to check for the specificity of *CHOP* primers. As expected, *CHOP* v1 band was only detected using the *CHOP* v1 plasmid template at the size of 214 bp, whereas *CHOP* v2 was found only with the *CHOP* v2 plasmid template at 207 bp in the agarose gel electrophoresis (Figure 5B).

We next used the *CHOP* v1 and *CHOP* v2 primer sets in the PCR analysis to detect whether variants were in liver tissues of mice treated via IP with vehicle or tunicamycin, another ER stressor that blocks N-glycosylation. *CHOP* v1 DNA band was amplified in the Tm-treated liver template at the size of 214 bp, whereas *CHOP* v2 band was detected in the Tm-treated liver at the size of 207 bp. In comparison to *CHOP* v1, the diminished intensity of *CHOP* v2 band suggests low expression of *CHOP* v2 transcript in liver tissues (Figure 5C).

CHOP **transcripts are preferentially translated during cellular stresses**

The 5'-leader of *CHOP* v1 contains an uORF, applying an inhibitory effect on *CHOP* mRNA translation in non-stressed conditions. Repression of *CHOP* translation is overcome by eIF2α phosphorylation in response to diverse stresses. Preferential translation of *CHOP* is facilitated by a mechanism involving ribosomal bypass of the inhibitory effect of *CHOP* v1 uORF (Palam et al., 2011). To address whether the alternative splicing in the 5'-leader of *CHOP* can affect the regulation of *CHOP* translation, we transiently transfected a firefly luciferase reporter encoding either the 5'-leader sequences of *CHOP* v1 or *CHOP* v2. In each reporter construct, the initiation codon of the *CHOP* was fused in-frame with the luciferase coding sequence. Both reporters were transcriptional expressed using a constitutive TK promoter. The CHOP v1-Luc and CHOP-Luc reporters were transiently transfected in wild-type MEF cells, in conjunction with a *Renilla* luciferase plasmid for normalization. The transfected cells were then cultured in the presence or absence of thapsigargin, a potent inducer of ER stress. We found more than a 2-fold induction of luciferase activity expressed from the CHOP-luciferase reporters, consistent with preferential translation of *CHOP* upon induced eIF2α phosphorylation and stress (Figure 6).

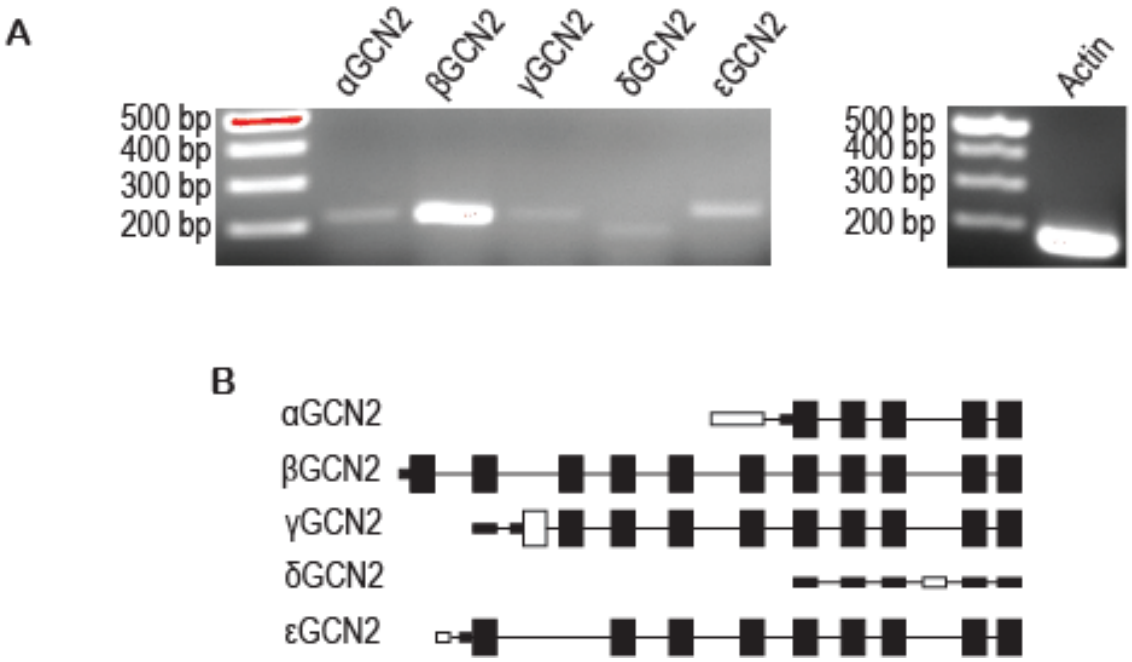

Figure 3: Semi-quantitative PCR analysis of *GCN2* variants in non-treated MEF cells and their different exon arrangement in the 5'-portion of the *GCN2* gene. (A) *GCN2* variants, including alpha (α*GCN2*), beta (β*GCN2*), gamma (γ*GCN2*), delta (δ*GCN2*), and epsilon (ε*GCN2*) transcripts, were detected in non-treated MEF cells at the expected sizes; α*GCN2* at 211 base pairs (bp), β*GCN2* at 211 bp, γ*GCN2* at 212, δ*GCN2* at 172, and ε*GCN2* at 214 bp. The absence of larger bands supports specificity of the PCR primers. Also *β-actin* mRNA was measured in the MEF cell preparations as a control. (B) β*GCN2* is the full-length canonical sequence and α*GCN2* mRNA contains a large exon that is either shortened or absent in other isoforms. Both γ*GCN2* and δ*GCN2* mRNAs have an additional exon that is spliced out in the other variants. In comparison to γ*GCN2* and δ*GCN2*, ε*GCN2* mRNA has also a spliced exon corresponding to the third exon of β*GCN2* transcript. The common exons are represented as black boxes, whereas the white boxes indicate the exons unique to the spliced variant; the small white and/or black boxes represent the 5'-UTR of *GCN2* isoforms; introns are illustrated as lines connecting exons designated as boxes.

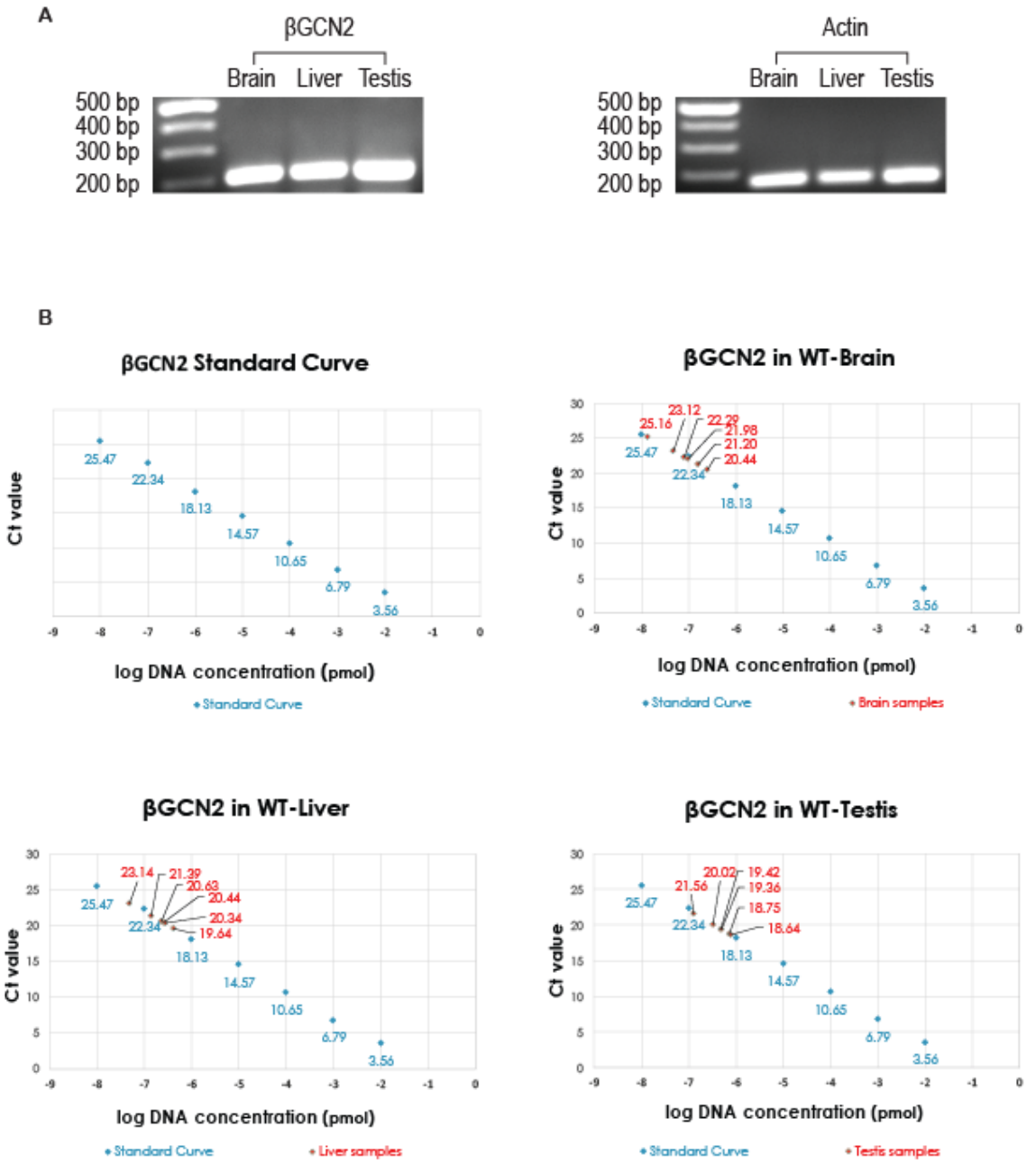

Figure 4: Expression of the beta isoform of *GCN2* (β*GCN2*) and measurement of β*GCN2* in different mouse tissues. (A) β*GCN2* mRNA is expressed in wild type (WT)-untreated (NT)-whole mouse brain, liver and testis. Also, *β-actin* mRNA was detected in the indicated tissues as an internal control. (B) The β*GCN2* standard curve is created by the quantitative RT-PCR of serial dilutions of β*GCN2* DNA. This standard curve can be used to determine how many β*GCN2* transcripts in unknown samples such as those derived from the whole mouse

brain, liver, and testis. β*GCN2* transcripts are estimated in the different mouse tissues by plotting the Y variable of the slope of β*GCN2* standard curve with the values of cycle threshold (Ct) of unknown samples (Y = - 3.7254X - 4.1254). Thus, the X variable, which represents the DNA concentration of β*GCN2*, can be calculated. According to the results of RT-PCR, β*GCN2* expression is high in testis, moderate in liver and low in brain.

Table 4: RT-PCR result using different sample concentrations from different mouse tissues and the calculated B*GCN2* transcripts in each sample.

Sample concentrations (µg)	WT-brain		WT-liver		WT-testis	
	Ct values	β*GCN2* amount	Ct values	β*GCN2* amount	Ct values	β*GCN2* amount
0.2	25.16	$1.4x10^{-08}$	23.14	$4.8x10^{-08}$	21.56	$1.3x10^{-07}$
0.4	23.12	$4.9x10^{-08}$	21.39	$1.4x10^{-07}$	20.02	$3.3x10^{-07}$
0.6	22.29	$8.1x10^{-08}$	20.63	$2.3x10^{-07}$	19.42	$4.8x10^{-07}$
0.8	21.98	$9.8x10^{-08}$	20.44	$2.6x10^{-07}$	19.36	$5x10^{-07}$
1	21.20	$1.6x10^{-07}$	20.34	$2.7x10^{-07}$	18.75	$7.2x10^{-07}$
1.2	20.44	$2.6x10^{-07}$	19.64	$4.2x10^{-07}$	18.64	$7.8x10^{-07}$

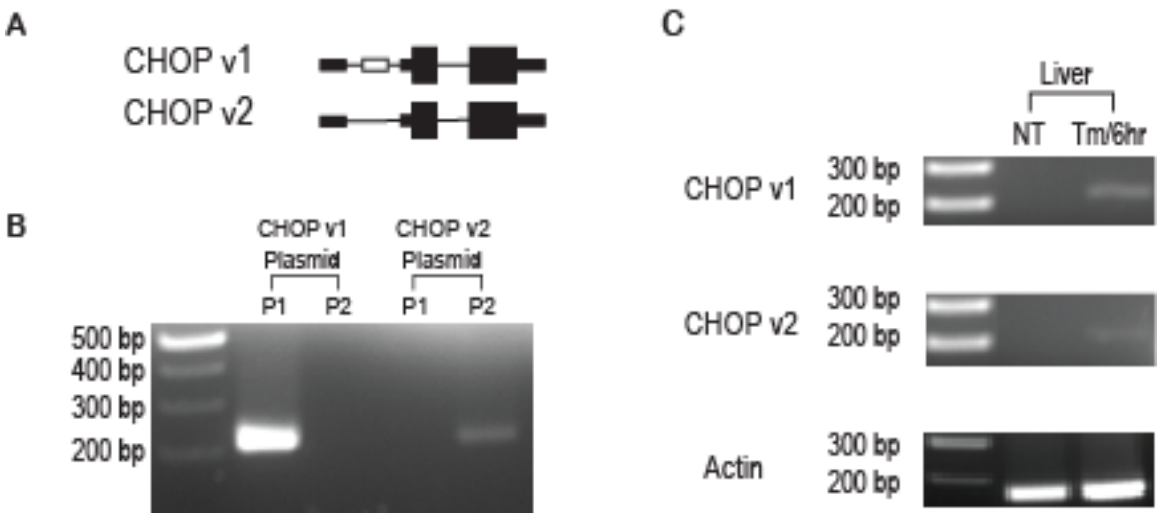

Figure 5: Semi-quantitative PCR analysis of *CHOP* splicing variants in liver tissues. (A) *CHOP* v1 mRNA contains four exons, whereas *CHOP* v2 mRNA lacks the second exon of *CHOP* v1. The 5'-leader of *CHOP* is encoded within the first and second exons, while the coding sequence is encoded within the third and fourth exons of *CHOP*. (B) As expected, *CHOP* v1 primers amplified *CHOP* v1 transcripts at the size of 214 bp in the *CHOP* v1 plasmid template containing the full sequence of *CHOP* v1. While, *CHOP* v2 isoforms were detected at the size of 207 bp in the *CHOP* v2 plasmid expressing the whole sequence of *CHOP* v2. The presence of a single band amplified by the corresponding primers in each plasmid indicates the specificity of *CHOP* primers. (C) *CHOP* v1 and *CHOP* v2 are also expressed in tunicamycin-treated liver tissues at the expected sizes 214 bp and 207 bp, respectively. Also, *β-actin* was used as a control. The common exons are represented as black boxes, and the alternative exon is shown as a white box; small white and/or black boxes represent either the 5'- or 3'- UTRs of *CHOP* transcripts; introns are illustrated as lines connecting exons; P1, primers specific for *CHOP* v1 detection; P2, primers specific for *CHOP* v2 detection; NT, non-treated liver tissues; Tm/6hr, liver tissues treated with tunicamycin for 6 hours.

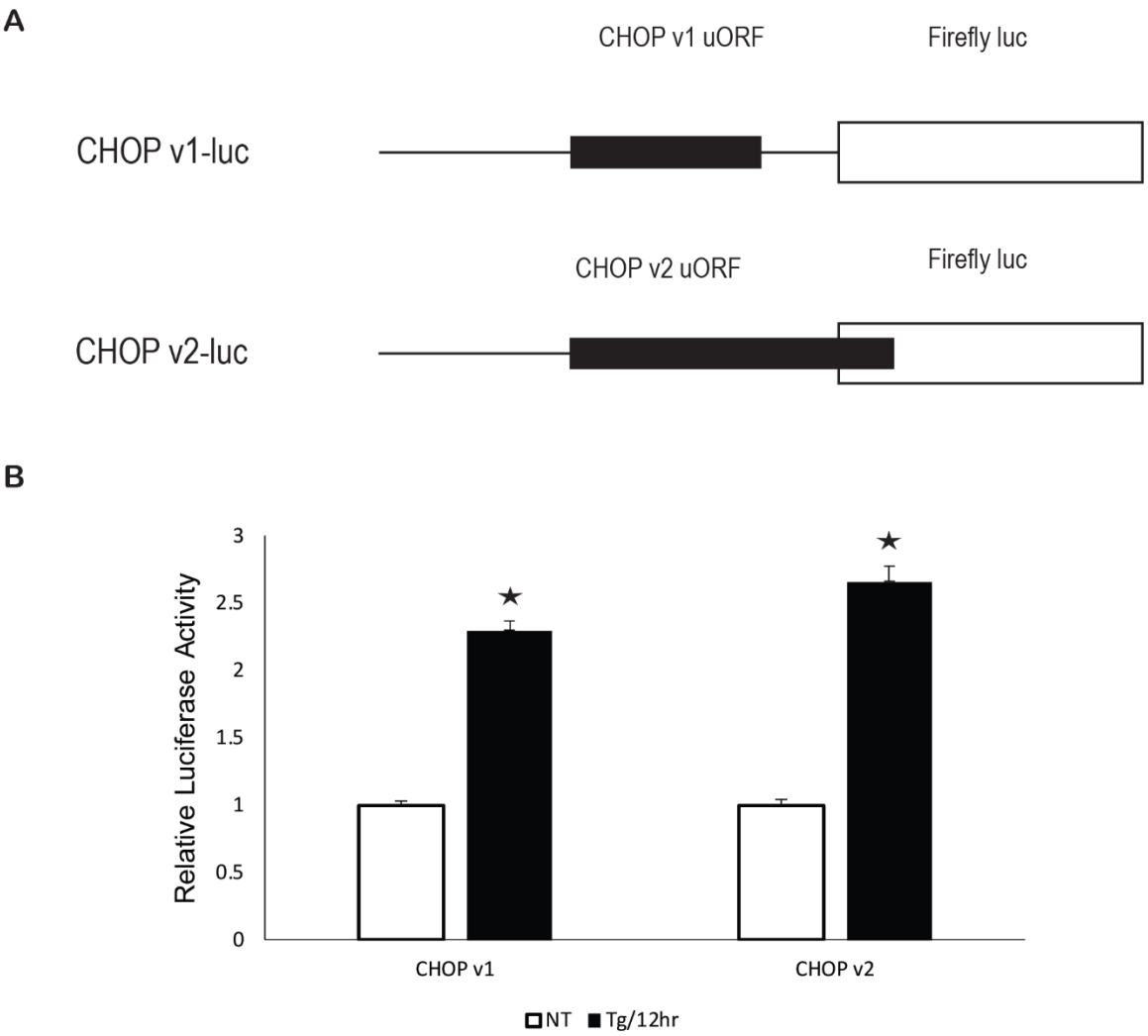

Figure 6: Representation of the two *CHOP* constructs (CHOP v1-luc and CHOP v2-luc), and their luciferase activity in normal and stressed conditions. (A) The CHOP v1-luc plasmid is constructed to have an uORF preceding the *luc* ORF to determine the effect of *CHOP* v1 uORF on luciferase activity. The CHOP v2-luc plasmid is constructed to have *CHOP* v2 uORF that overlaps out-of-frame with the *luciferase* CDS. (B) In comparison to non-treated (NT) MEF cells, *CHOP* v1 and *CHOP* v2 showed nearly a 2.5-fold induction of luciferase activity in MEF cells treated with thapsigargin (Tg), an ER stress agent, for 12 hours (12 hr). These findings indicate that *CHOP* v1 and *CHOP* v2 are preferentially translated during ER stress.

Discussion

The mammalian eIF2α kinases respond to alterations in the cell homeostasis initiated by environmental and physiological stresses. For example, the protein kinase GCN2 is a multidomain protein that phosphorylates Ser-51 of eIF2α in response to nutrient starvation. GCN2 senses the deficiency of amino acids through binding of uncharged tRNAs to the HisRS domain of GCN2. This binding elicits a conformational change that activates the cKD leading to high levels of eIF2α phosphorylation (Castilho et al., 2014).

mRNA diversity is driven by regulatory mechanisms including alternative splicing and/or alternative utilization of promotors. As a result, transcripts acquire different N-terminal sequences or 5'UTRs among transcripts (Landry, Mager, & Wilhelm, 2003). For example, α*GCN2*, β*GCN2*, γ*GCN2* have been proposed to be a consequence of alternative splice sites and different promoters that lead to different transcriptional start sites (Sood et al., 2000; Zhang et al., 2002). Similarly, the new isoforms, δ*GCN2*, and ε*GCN2,* likely arise from different promoters and alternative splicing sites. In addition to alternative usage of promoters and alternative splicing, *GCN2* may be subject to alternative polyadenylation resulting in different 3'-UTRs encoded in α*GCN2* and γ*GCN2* transcripts. The 3'-ends of mRNAs are known to be targeted by microRNAs, short non-coding RNAs, that negatively regulate gene expression (van Kouwenhove, Kedde, & Agami, 2011).

Mouse *GCN2* is suggested to produce five transcripts, α*GCN2*, β*GCN2*, γ*GCN2*, δ*GCN2*, and ε*GCN2*, which differ in the encoded N-terminal polypeptide sequences (Figure 3). The N-terminal domain of GCN2 is essential for GCN1 binding in yeast (Sattlegger & Hinnebusch, 2000). The conservation of key residues within the NTD among GCN2 orthologues implies a similar mechanism of function (Nameki et al., 2004). These findings suggest that the activities of all *GCN2* isoforms are not uniformly functional or regulated in a similar fashion. For example, δ*GCN2* encodes a shorter protein because it lacks the NTD and pKD. The shorter version of GCN2 protein may resemble the function of the truncated protein kinase of baculovirus that dampens the kinase activity of human and

yeast eIF2α kinases (Dever et al., 1998). However, the negative feedback of the δ*GCN2* on eIF2α kinases awaits further biochemical analysis.

To evaluate the differential expression of *GCN2* transcripts, RT-PCR results for β*GCN2* revealed that β*GCN2* transcripts are more abundant in the testis tissue than in the liver tissue, while the brain tissue shows the least expression of β*GCN2* in non-stressed conditions. In the future, two questions will be worth pursuing: first, which *GCN2* isoform is dominant during stress? Second, which tissues would the stress-dominant *GCN2* transcript be most highly expressed? Detection of β*GCN2* in different mouse tissues is consistent with those reported (Berlanga et al., 1999; Sood et al., 2000). The absence of the two additional *GCN2* variants, δ*GCN2* and ε*GCN2*, in mouse brain, liver, and testis tissues suggests that each is minimally expressed or restricted to other tissues. If δ*GCN2* and ε*GCN2* are tissue-specific transcripts, this will open new avenues for diagnostic biomarkers or therapeutic targets.

CHOP is another ISR key regulator that determines the cell fate during stress. As a result of eIF2α phosphorylation in stressed cells, the cells produce ATF4 proteins in an attempt to restore cellular homeostasis; alternatively, ATF4 protein promotes *CHOP* transcriptional expression during prolonged stressed conditions. Consequent to increased eIF2α phosphorylation, accumulation of *CHOP* mRNA induces an apoptotic pathway leading to a program of gene expression that triggers cellular death.

CHOP mRNA is alternatively spliced at the second exon into *CHOP* v1 and *CHOP* v2 variants. Both *CHOP* transcripts encode the full-length protein, but differ in their 5'-portions that carry out a regulatory role in *CHOP* translation. *CHOP* v1 transcript appears to be more highly expressed compared to *CHOP* v2 in tunicamycin-treated liver tissues (Figure 5C). This finding may indicate that there is differential expression of *CHOP* transcripts among mouse tissues.

To address the effect of the different 5'-leaders of *CHOP* on the regulation of *CHOP* translation, we developed the dual luciferase assays for CHOP v1-luc and CHOP v2-luc constructs (Figure 6). Consistent with previous results (Palam et al., 2011), *CHOP* v1 uORF functions to prevent translation at the downstream

coding sequence of *CHOP* mRNA in non-stressed cells. The inhibition of *CHOP* translation has been attributed to a specific sequence within *CHOP* v1 uORF that is conserved among eukaryotes. The uORF sequence, encoding IFIHHHT residues, has been suggested to cause a ribosomal stall blocking other elongating ribosomes, and thus contributes to reduced CHOP proteins in non-stressed conditions (Young et al., 2016). However, CHOP proteins under stressed conditions, are synthesized by the ribosomal bypass model. In the bypass model, the scanning ribosomes ignore the initiation codons of *CHOP* v1 uORF in part because of the poor Kozak of its initiation codons; instead, the ribosomes recognize the downstream initiation codon of *CHOP* CDS and direct translation of the *CHOP* CDS (Palam et al., 2011).

Likewise, the *CHOP* v2 uORF is suggested to control *CHOP* translation in a similar fashion to the *CHOP* v1 uORF as illustrated in the luciferase reporter assay (Figure 6). Although *CHOP* v2 uORF is suggested to confer regulation of *CHOP* translation as described for *CHOP* v1, the *CHOP* v2 uORF lacks the inhibitory sequence, i.e. IFIHHHT, positioned in the *CHOP* v1 uORF. Therefore, the notion of regulated CHOP protein synthesis relies on another mechanism for *CHOP* v2 mRNA. The *CHOP* v2 uORF overlaps out-of-frame with the *CHOP* CDS, suggesting that the uORF serves as an inhibitor to translation by allowing translating ribosomes to elongate past the initiation codon of the *CHOP* CDS. After translation of the *CHOP* v2 uORF, ribosomes would be 3’ of the of the CDS initiation codon and require substantial 3’ to 5’ scanning, which is not thought to occur. It is suggested that the *CHOP V1* mRNA is subject to nonsense mediated decay, or similar mRNA surveillance systems that may be triggered by stalled ribosomes in the *CHOP* v1 uORF. The apparent absence of stalled ribosomes in the *CHOP* v2 uORF may indicate differences in mRNA turnover, which could be important for *CHOP* expression late in the progression of the stress responses. Differences between *CHOP* v1 and *CHOP* v2 mRNA decay during ER stress responses warrants future investigation.

References

Baird, T. D., & Wek, R. C. (2012). Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism. *Adv Nutr, 3*(3), 307-321. doi:10.3945/an.112.002113

Banks, A. S., Kim-Muller, J. Y., Mastracci, T. L., Kofler, N. M., Qiang, L., Haeusler, R. A., . . . Accili, D. (2011). Dissociation of the glucose and lipid regulatory functions of FoxO1 by targeted knockin of acetylation-defective alleles in mice. *Cell Metab, 14*(5), 587-597. doi:10.1016/j.cmet.2011.09.012

Berlanga, J. J., Santoyo, J., & De Haro, C. (1999). Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase. *Eur J Biochem, 265*(2), 754-762.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem, 72*, 291-336. doi:10.1146/annurev.biochem.72.121801.161720

Castilho, B. A., Shanmugam, R., Silva, R. C., Ramesh, R., Himme, B. M., & Sattlegger, E. (2014). Keeping the eIF2 alpha kinase Gcn2 in check. *Biochim Biophys Acta, 1843*(9), 1948-1968. doi:10.1016/j.bbamcr.2014.04.006

Chen, J. J. (2007). Regulation of protein synthesis by the heme-regulated eIF2alpha kinase: relevance to anemias. *Blood, 109*(7), 2693-2699. doi:10.1182/blood-2006-08-041830

Costa-Mattioli, M., Gobert, D., Harding, H., Herdy, B., Azzi, M., Bruno, M., . . . Sonenberg, N. (2005). Translational control of hippocampal synaptic plasticity and memory by the eIF2alpha kinase GCN2. *Nature, 436*(7054), 1166-1173. doi:10.1038/nature03897

Delepine, M., Nicolino, M., Barrett, T., Golamaully, M., Lathrop, G. M., & Julier, C. (2000). EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nat Genet, 25*(4), 406-409. doi:10.1038/78085

Dever, T. E., Sripriya, R., McLachlin, J. R., Lu, J., Fabian, J. R., Kimball, S. R., & Miller, L. K. (1998). Disruption of cellular translational control by a viral truncated eukaryotic translation initiation factor 2alpha kinase homolog. *Proc Natl Acad Sci U S A, 95*(8), 4164-4169.

Eyries, M., Montani, D., Girerd, B., Perret, C., Leroy, A., Lonjou, C., . . . Soubrier, F. (2014). EIF2AK4 mutations cause pulmonary veno-occlusive disease, a recessive form of pulmonary hypertension. *Nat Genet, 46*(1), 65-69. doi:10.1038/ng.2844

Fusakio, M. E., Willy, J. A., Wang, Y., Mirek, E. T., Al Baghdadi, R. J., Adams, C. M., . . . Wek, R. C. (2016). Transcription factor ATF4 directs basal and stress-induced gene expression in the unfolded protein response and cholesterol metabolism in the liver. *Mol Biol Cell, 27*(9), 1536-1551. doi:10.1091/mbc.E16-01-0039

Garcia-Barrio, M., Dong, J., Ufano, S., & Hinnebusch, A. G. (2000). Association of GCN1-GCN20 regulatory complex with the N-terminus of eIF2alpha

kinase GCN2 is required for GCN2 activation. *EMBO J, 19*(8), 1887-1899. doi:10.1093/emboj/19.8.1887

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., . . . Ast, G. (2006). Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell, 22*(6), 769-781. doi:10.1016/j.molcel.2006.05.008

Guo, F., & Cavener, D. R. (2007). The GCN2 eIF2alpha kinase regulates fatty-acid homeostasis in the liver during deprivation of an essential amino acid. *Cell Metab, 5*(2), 103-114. doi:10.1016/j.cmet.2007.01.001

Habjan, M., Pichlmair, A., Elliott, R. M., Overby, A. K., Glatter, T., Gstaiger, M., . . . Weber, F. (2009). NSs protein of rift valley fever virus induces the specific degradation of the double-stranded RNA-dependent protein kinase. *J Virol, 83*(9), 4365-4375. doi:10.1128/JVI.02148-08

Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., & Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell, 6*(5), 1099-1108.

Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calfon, M., . . . Ron, D. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell, 11*(3), 619-633.

Hinnebusch, A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol, 59*, 407-450. doi:10.1146/annurev.micro.59.031805.133833

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol, 11*(2), 113-127. doi:10.1038/nrm2838

Jiang, H. Y., & Wek, R. C. (2005). GCN2 phosphorylation of eIF2alpha activates NF-kappaB in response to UV irradiation. *Biochem J, 385*(Pt 2), 371-380. doi:10.1042/BJ20041164

Kalsotra, A., & Cooper, T. A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet, 12*(10), 715-729. doi:10.1038/nrg3052

Kwon, N. H., Kang, T., Lee, J. Y., Kim, H. H., Kim, H. R., Hong, J., . . . Kim, S. (2011). Dual role of methionyl-tRNA synthetase in the regulation of translation and tumor suppressor activity of aminoacyl-tRNA synthetase-interacting multifunctional protein-3. *Proc Natl Acad Sci U S A, 108*(49), 19635-19640. doi:10.1073/pnas.1103922108

Lageix, S., Rothenburg, S., Dever, T. E., & Hinnebusch, A. G. (2014). Enhanced interaction between pseudokinase and kinase domains in Gcn2 stimulates eIF2alpha phosphorylation in starved cells. *PLoS Genet, 10*(5), e1004326. doi:10.1371/journal.pgen.1004326

Lageix, S., Zhang, J., Rothenburg, S., & Hinnebusch, A. G. (2015). Interaction between the tRNA-binding and C-terminal domains of Yeast Gcn2 regulates kinase activity in vivo. *PLoS Genet, 11*(2), e1004991. doi:10.1371/journal.pgen.1004991

Landry, J. R., Mager, D. L., & Wilhelm, B. T. (2003). Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet, 19*(11), 640-648. doi:10.1016/j.tig.2003.09.014

Masuoka, H. C., & Townes, T. M. (2002). Targeted disruption of the activating transcription factor 4 gene results in severe fetal anemia in mice. *Blood, 99*(3), 736-745.

McCullough, K. D., Martindale, J. L., Klotz, L. O., Aw, T. Y., & Holbrook, N. J. (2001). Gadd153 sensitizes cells to endoplasmic reticulum stress by down-regulating Bcl2 and perturbing the cellular redox state. *Mol Cell Biol, 21*(4), 1249-1259. doi:10.1128/MCB.21.4.1249-1259.2001

McIlwain, D. R., Berger, T., & Mak, T. W. (2013). Caspase functions in cell death and disease. *Cold Spring Harb Perspect Biol, 5*(4), a008656. doi:10.1101/cshperspect.a008656

Nameki, N., Yoneyama, M., Koshiba, S., Tochio, N., Inoue, M., Seki, E., . . . Yokoyama, S. (2004). Solution structure of the RWD domain of the mouse GCN2 protein. *Protein Sci, 13*(8), 2089-2100. doi:10.1110/ps.04751804

Palam, L. R., Baird, T. D., & Wek, R. C. (2011). Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *J Biol Chem, 286*(13), 10939-10949. doi:10.1074/jbc.M110.216093

Pang, Q., Christianson, T. A., Keeble, W., Koretsky, T., & Bagby, G. C. (2002). The anti-apoptotic function of Hsp70 in the interferon-inducible double-stranded RNA-dependent protein kinase-mediated death signaling pathway requires the Fanconi anemia protein, FANCC. *J Biol Chem, 277*(51), 49638-49643. doi:10.1074/jbc.M209386200

Puthalakath, H., O'Reilly, L. A., Gunn, P., Lee, L., Kelly, P. N., Huntington, N. D., . . . Strasser, A. (2007). ER stress triggers apoptosis by activating BH3-only protein Bim. *Cell, 129*(7), 1337-1349. doi:10.1016/j.cell.2007.04.027

Qiu, H., Dong, J., Hu, C., Francklyn, C. S., & Hinnebusch, A. G. (2001). The tRNA-binding moiety in GCN2 contains a dimerization domain that interacts with the kinase domain and is required for tRNA binding and kinase activation. *EMBO J, 20*(6), 1425-1438. doi:10.1093/emboj/20.6.1425

Romano, P. R., Garcia-Barrio, M. T., Zhang, X., Wang, Q., Taylor, D. R., Zhang, F., . . . Hinnebusch, A. G. (1998). Autophosphorylation in the activation loop is required for full kinase activity in vivo of human and yeast eukaryotic initiation factor 2alpha kinases PKR and GCN2. *Mol Cell Biol, 18*(4), 2282-2297.

Ron, D., & Habener, J. F. (1992). CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. *Genes Dev, 6*(3), 439-453.

Sattlegger, E., & Hinnebusch, A. G. (2000). Separate domains in GCN1 for binding protein kinase GCN2 and ribosomes are required for GCN2 activation in amino acid-starved cells. *EMBO J, 19*(23), 6622-6633. doi:10.1093/emboj/19.23.6622

Shao, J., Grammatikakis, N., Scroggins, B. T., Uma, S., Huang, W., Chen, J. J., . . . Matts, R. L. (2001). Hsp90 regulates p50(cdc37) function during the biogenesis of the activeconformation of the heme-regulated eIF2 alpha kinase. *J Biol Chem, 276*(1), 206-214. doi:10.1074/jbc.M007583200

Sood, R., Porter, A. C., Olsen, D. A., Cavener, D. R., & Wek, R. C. (2000). A mammalian homologue of GCN2 protein kinase important for translational control by phosphorylation of eukaryotic initiation factor-2alpha. *Genetics, 154*(2), 787-801.

Staley, J. P., & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell, 92*(3), 315-326.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., . . . Soreq, H. (2005). Function of alternative splicing. *Gene, 344*, 1-20. doi:10.1016/j.gene.2004.10.022

Tabas, I., & Ron, D. (2011). Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nat Cell Biol, 13*(3), 184-190. doi:10.1038/ncb0311-184

Tanaka, T., Tsujimura, T., Takeda, K., Sugihara, A., Maekawa, A., Terada, N., . . . Akira, S. (1998). Targeted disruption of ATF4 discloses its essential role in the formation of eye lens fibres. *Genes Cells, 3*(12), 801-810.

van Kouwenhove, M., Kedde, M., & Agami, R. (2011). MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer, 11*(9), 644-656. doi:10.1038/nrc3107

Vattem, K. M., & Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A, 101*(31), 11269-11274. doi:10.1073/pnas.0400541101

Wek, R. C., Jackson, B. M., & Hinnebusch, A. G. (1989). Juxtaposition of domains homologous to protein kinases and histidyl-tRNA synthetases in GCN2 protein suggests a mechanism for coupling GCN4 expression to amino acid availability. *Proc Natl Acad Sci U S A, 86*(12), 4579-4583.

Xu, L., Su, L., & Liu, X. (2012). PKCdelta regulates death receptor 5 expression induced by PS-341 through ATF4-ATF3/CHOP axis in human lung cancer cells. *Mol Cancer Ther, 11*(10), 2174-2182. doi:10.1158/1535-7163.MCT-12-0602

Yamaguchi, H., & Wang, H. G. (2004). CHOP is involved in endoplasmic reticulum stress-induced apoptosis by enhancing DR5 expression in human carcinoma cells. *J Biol Chem, 279*(44), 45495-45502. doi:10.1074/jbc.M406933200

Yang, R., Wek, S. A., & Wek, R. C. (2000). Glucose limitation induces GCN4 translation by activation of Gcn2 protein kinase. *Mol Cell Biol, 20*(8), 2706-2717.

Young, S. K., Palam, L. R., Wu, C., Sachs, M. S., & Wek, R. C. (2016). Ribosome Elongation Stall Directs Gene-specific Translation in the Integrated Stress Response. *J Biol Chem, 291*(12), 6546-6558. doi:10.1074/jbc.M115.705640

Zaborske, J. M., Narasimhan, J., Jiang, L., Wek, S. A., Dittmar, K. A., Freimoser, F., . . . Wek, R. C. (2009). Genome-wide analysis of tRNA charging and

activation of the eIF2 kinase Gcn2p. *J Biol Chem, 284*(37), 25254-25267. doi:10.1074/jbc.M109.000877

Zhang, P., McGrath, B. C., Reinert, J., Olsen, D. S., Lei, L., Gill, S., . . . Cavener, D. R. (2002). The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. *Mol Cell Biol, 22*(19), 6681-6688.

Curriculum Vitae

Mohammed Alzahrani

Objective

To have the opportunity to pursue a Doctorate's Degree in the field of Biochemistry, so that I can pass the knowledge to the next generation while serving the world of science.

Education Attainment

2005 – 2010	Bachelor's Degree in Laboratory Medicine with an overall grade excellent (89.14%) from the Umm Al-Qura University (UQU)-Faculty of Applied Medical Sciences, Makkah, Saudi Arabia (SA).
2014-2016	Master of Science in Biochemistry and Molecular Biology from the Indiana University (IU), Indianapolis, Indiana, United States of America (USA).

Research Experience

In 2009, my colleagues and I conducted a graduation research at UQU centered on the topic of "The Prevalence of Anti-Hepatitis E Virus Antibodies among Male Blood Donors." The research was under the supervision of Dr. Ayman Johargy in 2009 in SA. This project aimed to study the seroprevalence of hepatitis E virus in consented blood samples using enzyme-linked immunosorbent assay and determine some possible contributions such as participant's age, residence, and source of drinking water on the prevalence of anti-hepatitis E virus antibodies.

In 2016, the second project was entitled "The Effect of Alternative Splicing on Key Regulators of the Integrated Stress Response." This project was conducted in Dr. Ronald Wek's lab at the Indiana University-School of Medicine. The research studies how alternative splicing can influence the protein expression and function and distribution of *GCN2* and *CHOP* in different mouse

tissues. By using real time-polymerase chain reaction, we estimated the abundance of β*GCN2* transcript in the mouse brain, liver and testis tissues to identify which mouse tissue has the highest expression of β*GCN2*. Also, we detected two isoforms of *CHOP* in liver treated with tunciamycin for 6 hours that lead to endoplasmic reticulum stress. By employing a dual luciferase reporter assay, we determined the translation regulation of *CHOP* by the different 5'-ends of *CHOP* mRNAs suggested to be caused by alternative splicing.

Work Experience

2010-2011	Working as a Medical Technologist at the Hospital of King Abdulaziz Medical City-National Guard Health Affairs, SA.
2012	Working as a Teaching Assistant in the Biochemistry Department at King Saud bin Abdulaziz University for Health Sciences, SA.

Symposiums and Conferences

2010	6th Applied Medical Sciences Students Meeting for giving an Oral Presentation (12 hrs CME) and for Attendance (12 hrs CME).
2010	First Aid Course provided by New Future Training and Development (5 hrs CME).
2012	Basic Live Support provided by Saudi Council for Health Specialties (4 hrs CME).
2014-2015	Attending and presenting a scientific paper in the Central Dogma Journal Club of Molecular Biology at IU.

Communication Skills

- Arabic and English Languages (speaking and writing).
- A good experience in using internet for searching and sharing information.
- Using Microsoft Office and Adobe programs, e.g., Excel and Photoshop.